

92

1. The first step in the process is to identify the problem or issue that needs to be addressed. This involves gathering information and understanding the context of the problem.

(10) International Publication Number
WO 01/40967 A2

- WO 01/40967 A2**

BEST AVAILABLE COPY

Medical Record EclipsingArea of the Invention

The present invention relates to computer-based storage and intelligent retrieval of documents. In particular, the present invention relates to systems, methods and computer program products for transforming diverse document formats and computer systems to a selected, standardized format on a selected computer system.

Background of the Invention

10 A major obstacle in storage and handling of large amounts of information from different sources is incompatible document format. Historically integration of documents with different formatting was impractical because of the diversity of formats and systems used to process those formats. However, in recent years, there has been a growing agreement in the industry that standardized markup languages are well suited to address this kind of incompatibly structured information on 15 diverse computer systems. HTML has been developed for internet-based data and systems. This has been followed by the development of Standard General Markup Language (SGML), and a subset of that language eXtensible Markup Language (XML).

20 Nowadays it is practical to agree to exchange information in markup language formats based on some standard like SGML or XML. Where a consensus on data formats and systems cannot be reached, the use of SGML or XML allows each party to define its own formats and/or customize existing ones, while preserving the ability for others to read and process these files. In those cases where 25 it is not possible to agree on a format newly acquired information in a different format usually can be transformed into a canonical representation based on SGML or XML. Thus it is possible that all kinds of data from whatever source can be either received in an XML based format or can be transformed into XML by the recipient.

30 For a fuller discussion of SGML and XML see W3C. It is shorthand for *World Wide Web Consortium*, an international consortium of companies involved with the Internet and the Web. The W3C was founded in 1994 by Tim Berners-Lee,

p3

the original architect of the World Wide Web. The W3C is the chief standards body for HTTP and HTML. See also biztalk.org biztalk, a web site sponsored by Microsoft Corp. and relating to XML.

5 A common requirement for storing documents is to preserve the original file as received. This is needed in case a legal dispute arises concerning the contents of a document (for instance, whether out-of-normal lab data had really been available to a physician at a particular date). One approach would be to keep the original data in its original format but to transform or customize it to another version for local processing.

10 The ubiquitous usage of mark-up languages such as XML, and a special indexing system such as described in co-pending UK application titled " Virtual Indexing " and filed on even date herewith, allows one to use the original file format even during processing time. This eliminates the need to replicate the information into another data store.

15 It is quite useful to be able to combine data from different data sources for any number of reasons. In the example of aggregated health care data, possible users would be the patients themselves (legally enabled to control the data that is stored about them), physicians, insurance companies or researchers. Each case requires different security scenarios. For instance, the patient would only be allowed to
20 access data about himself or herself. A physician might have read/write access to the files of all her own patients but only read access to other physician's data. A researcher might be allowed to see all medical information, if it is anonymized (personal information would have to be suppressed).

Controlling the access to the contents of documents in a format defined by
25 other parties is no easy task. In particular, the format chosen by the party which defined the format might lend itself well to eclipsing sensitive data.

Assuming a suitably well defined document format, the transformation of the original source document into the desired result format can be implemented using a generic transformation language. For instance, in many cases it is possible to
30 transform simple ASCII text using regular expressions as implemented in the standard POSIX tool "sed" (stream editor).

For data expressed in a mark up language such as XML, finding a suitable transformation tool is straight forward as transforming XML into other formats (including different XML representations) is well understood. In particular, the W3C recommendation for XSLT defines a standard template language for expressing transformations of XML document.

The core of the present invention lies in the process of dynamically building transformation rules based on a specific (XML) document type and the user information.

Summary of the Invention

This invention relates to a process of dynamically building transformation rules based on a specific document type and the user information, the steps comprising:

1. requesting document D by U,
2. retrieving document D from database,
3. determining type T of document D,
4. generating transformation program P based on U and T,
5. executing P,
6. generating D', and
7. returning D' to U.

Description of the Figures

Figure 1 is a flowchart of the process of transforming data.

Description of the Invention

The present invention finds particular, but not exclusive, application to the healthcare industry. A problem frequently encountered in healthcare systems is that medical information is generated in a variety of formats and styles from a multitude of data sources, such as physicians, labs, health care authorities and ultimately patients themselves. For all involved parties it is most valuable to access and analyze this data electronically for different purposes. A patient, for example, would be able to not only view his/her medical record, but accessing all prescription data or all data relating to a particular disease.

The core of the present invention lies in the process of dynamically building transformation rules based on a specific (XML) document type and the user information. The formal definition of the problem is:

85

Given the current user **U** and a request for a document **D**, generate a version **D'** of the document which is suitably filtered given the known access right of **U**.

This process can be broken down into the following steps (see Figure 1):

1. document **D** requested by **U**,
- 5 2. retrieve document **D** from database,
3. determine type **T** of document **D**,
4. generate transformation program **P** based on **U** and **T**,
5. execute **P**, generating **D'**,
6. return **D'** to **U**.

10 Example using XML and XSLT

The following example describes how this process can be applied to a discharge letter using XML or XSLT (XSL Transformations). Parenthetically, XSL (Extensible Style Language) is a specification for separating style from content when creating HTML or XML pages. The specifications work much like templates, allowing designers to apply single style documents to multiple pages. XSL is the second style specification to be offered by the World Wide Web Consortium (W3C) (www.w3c.org). The first, called Cascading Style Sheets (CSS), is similar to XSL but does not include two major XSL's innovations -- allowing developers to dictate the way Web pages are printed, and specifications allowing one to transfer XML documents across different applications. W3C released the first draft of XSL in August 1998, and promotes the specifications as helpful to the Web's speed, accessibility, and maintenance. A patient might have the right to read all information in discharge letters about himself, with the exception of subjective comments made by the physicians.

25 In this particular example, standard HTML mark up is used to define the document structure, and special mark up was added to mark special parts of the content:

```

30  <discharge_letter patient_id="123">
    <html:p>
        The patient complained about <med:symptom>fever</med:symptom>. The
        result of the examination was that it was in fact just a
        <med:diagnosis>regular cold</med:diagnosis>.
        <med:subjective_comment>His parents should take care about how he
        dresses during winter.</med:subjective_comment>
35    </html:p>
    </discharge_letter>

```

p6

To filter this letter based for display to the patient, the contents of the XML tag "med:subjective_comment" needs to be removed. Below is a XSLT program which would be generated for a patient with the internal identification number 123. It would suppress the whole document for those documents where the patient identification does not match. For those where it does match, it will still remove the contents of the aforementioned tag.

```

<xsl:stylesheet xmlns:xsl="http://www.w3.org/TR/WD-xsl">

10      <!-- copy element and recurse -->
      <xsl:template>
      <xsl:copy>
      <xsl:apply-templates select="@* | * | comment() | pi() | text()"/>
      </xsl:copy>
15      </xsl:template>

      <xsl:template match="med:subjective_comment">
      <!-- suppress contents of this element -->
      </xsl:template>

20      <xsl:template match="/discharge_letter[@patid='123']">
      <xsl:apply-templates />
      </xsl:template>

25      <xsl:template match="/discharge_letter[@patid!='123']">
      <!-- empty -->
      </xsl:template>

      </xsl:stylesheet>

```

30 The foregoing illustrations are given to exemplify the invention, not to limit it. Reference is made to the claims for what is reserved to the inventors hereunder.

P7

What is claimed is:

1. In a computational system, a process of dynamically building transformation rules based on a specific document type and the user information, the
5 steps comprising:
 8. requesting document D by U,
 9. retrieving document D from database,
 10. determining type T of document D,
 11. generating transformation program P based on U and T,
 - 10 12. executing P,
 13. generating D', and
 14. returning D' to U.
2. The process of claim 1 wherein the transformation is based on a standard markup language which is HTML, SGML, XML or XSLT.

98

1/1

Figure 1

